

Short Paper #1: Ethics of "Taste, Ties, and Time"

Ke Wang

1 Overview

At first glance, the "Taste, Ties, and Time" project is very valuable given its completeness, richness, naturalness and many other features. However, a close analysis by Salganik's four principles of ethical research reveals many ethical landmines in the goldmine of digital data.

2 Ethical analysis

2.1 Respect for persons

The T3 project violates this principle, which requires treating people as autonomous and considering their wishes. Firstly, the T3 researchers did not ask consent form from participants. Instead, they directly downloaded and shared the profiles. They did get approval from IRB and Facebook company, but these alone are not sufficient. Secondly, there are reasons to assume that researchers had violated some participants' wishes, though participants were not directly asked. Since data were acquired by research assistants who used "in network" of Harvard College, those contents that were set as "only seen to the in network" were not intended to be shared with the public. What's more, participants may insist that their data only appear on facebook, not other places where they can not control the flow of their personal information. Though no real harm is caused, researchers violates dignity-based theory of privacy (Bloustein 1964). Even worse, the data contain students' housing information and personal email addresses provided by the university. These data were not from Facebook, so they should not be available to the public. Thirdly, releasing the data to the public invites many secondary uses of the data, again, without authorization by the students. Fourthly, subjects were not provided any access to view the data to correct for errors or unwanted information.

2.2 Beneficence

The T3 project could do better on this principle, which emphasizes maximizing possible benefits and minimizing possible harms. Admittedly, it did well in maximizing benefits, as its advantages (naturally occurring, sociocentric, multiplex, longitudinal, rich information) bring fruitful insights, including comparing mechanisms of the formation of friendship (Wimmer, & Lewis, 2010) and isolating social influence from social selection in similarity (Lewis, Gonzalez, & Kaufman, 2012). However, more steps should be done to minimize potential harm.

A major concern is that protection for privacy is not considerate enough, leaving the risk of re-indentification high. Though researchers stripped all names, the remaining rich information, which contain economic, cultural or social identity, still makes re-indentification highly possible. Researchers are aware of the sensitivity of the cultural fingerprint in cultural taste data, but their choice of delaying in release of cultural taste data only delayed the possible re-indentification without reducing its possibility. Considering the example given by Zimmer (2010), some states has only one single student represented in the dataset. Combining external resouces like local news, it is easy to identify the single student. Once a student is identified, other students in his/her social network on facebook will also be easily identified. Furermore, they did not screen out participants who are vulnerable.

2.3 Justice

The principle of Justice argues that risks and benefits of research should be distributed fairly across groups.

If sticking strictly to this principle, researchers should collect samples randomly from the facebook community instead of using only students in Harvard College, who are not representative of the population. Chances are that Harvard admission may involve selective bias based on race, gender or other categories. On the other hand, the sample has advantages than random sampling. First, researchers can combine data provided by the college, including data on roommates and personal email addresses to increase richness and convenience. Second, the sociocentric feature is good for certain research questions.

I think the principle of justice should be applied to a domain of research instead of individual projects. For example, if one project essentially concerns the effect of certain drug on black people, it is unreasonable to blame it for excluding sample of other population unless the field is filled with studies using black people as samples disproportionately. It is said that few network data have been collected from college students comparing to child or adolescent populations (Lewis et al., 2008), so the project does not violate this principle.

Finally, participants were not compensated financially. Researchers should ask participants whether they need compensation, as they “sell” their personal information and take informational risks.

2.4 Respect for law and public interest

Extending the principle of Beneficence, the principle of Respect for law and public interest requires two components: Compliance and transparency-based accountability.

As for compliance, it means that researchers should identify and obey relevant laws, contracts, and terms of service. The T3 project attempted to do so. It gained permission from facebook and the IRB of Harvard. After sharing the data to the public, requesters must sign a Terms and Conditions of Use statement. However, doubts exist that most people do not read the terms carefully (Gatt 2002), and T3 researchers may not be able to monitor or enforce compliance with these terms (Zimmer, 2010). I think it might help to change the form of “clicking through” the terms of conditions of use statement to “answering questions until pass”, ensuring understanding of important issues. In addition, it is unrealistic to expect researchers to monitor or enforce compliance of requesters. Maybe our society should build a credit system, and people should register with real identification to get access to the shared data.

For transparency-based accountability, it demands researchers take responsibility for their actions, to keep clearly aware of their goals, methods, and results at all stages of the research process. T3 project aligns with this principle. First, results are published in academic journals, enabling transparency-based accountability. Second, T3 researchers acted with responsibility. They exerted good-faith efforts to protect the privacy of the subjects (though they were overly optimistic), and withdrew data soon after they became aware of its high risk of identification.

3 My position

Though the features of the data are extremely tempting for me to analyze, I will not use it unless further actions are taken by the researchers to make it more ethical. Firstly, it is never too late for researchers ask participants for their consent. If some of them deny to provide or share their data publicly, researchers should delete them. They should also be provided access to view the data to correct for errors or unwanted information. Secondly, further steps should be processed to protect privacy. Given that it is known that the college is Harvard, the network fingerprint, cultural fingerprint and other identity information should be carefully treated. Maybe providing data of aggregate levels is better. Maybe it helps to check the background of requesters, and set stricter standard for access.

As researchers explore the goldmine of online data, many areas are still in the darkness. By learning from past lessons of other researchers, we should try to avoid ethical landmines so that we can conduct sustainable research in the digital world.

References

- Bloustein, E. (1964). Privacy as an aspect of human dignity: An answer to Dean Prosser. *New York University Law Review*, 39, 962–1007.
- Gatt, A. (2002). Click-wrap agreements the enforceability of click-wrap agreements. *Computer Law & Security Report*, 18(6), 404–410.
- Lewis, K., Gonzalez, M., & Kaufman, J. (2012). Social selection and peer influence in an online social network. *Proceedings of the National Academy of Sciences*, 109(1), 68-72.
- Lewis, K., Kaufman, J., Gonzalez, M., Wimmer, A., & Christakis, N. (2008). Tastes, ties, and time: A new social network dataset using Facebook.com. *Social networks*, 30(4), 330-342.
- Wimmer, A., & Lewis, K. (2010). Beyond and below racial homophily: Erg models of a friendship network documented on facebook. *American Journal of Sociology*, 116(2), 583-642.
- Zimmer, M. (2010). "But the data is already public": on the ethics of research in Facebook. *Ethics and information technology*, 12(4), 313-325.